DOCUMENT RESUME

ED 478 973 TM 035 118

AUTHOR	Jakwerth, Pamela M.; Stancavage, Frances B.
TITLE	An Investigation of Why Students Do Not Respond to Questions. NAEP Validity Studies. Working Paper Series.
INSTITUTION	National Center for Education Statistics (ED), Washington, DC.
REPORT NO	NCES-WP-2003-12
PUB DATE	2003-04-00
NOTE	27p.
AVAILABLE FROM	ED Pubs, P.O. Box 1398, Jessup, MD 20794-1398. Tel: 877-433-7827 (Toll Free); Web site: .
PUB TYPE	Reports - Research (143)
EDRS PRICE	EDRS Price MF01/PC02 Plus Postage.
DESCRIPTORS	Elementary Secondary Education; Interviews; *Junior High School Students; Junior High Schools; Responses; *Test Items; *Validity
IDENTIFIERS	*National Assessment of Educational Progress; *Omitted Responses

ABSTRACT

This study explored potential reasons behind student omission of responses to assessment questions. Understanding why students fail to answer certain questions may help inform the proper treatment of missing data during the estimation of item parameters and achievement distributions. The study was exploratory, small in scope, and qualitative in nature. The general approach was to visit schools in which the 1998 eighth-grade National Assessment of Educational Progress (NAEP) assessments in reading and civics were being conducted and interview samples of students about their test taking behaviors and their reasons for not answering particular questions following the assessment sessions. Interviews also attempted to determine whether the students could have answered the questions they left blank correctly. Eighty-four students were originally interviewed for this investigation, and 19 of these had omitted no questions. In previous studies, response rates appeared to vary with student characteristics, but it was not possible to draw statistically meaningful conclusions about the demographic characteristics to students likely to omit questions based on the makeup of the sample. (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

TM TMTR 0155

ED 478 973

NATIONAL CENTER FOR EDUCATION STATISTICS

Working Paper Series

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

NAEP VALIDITY STUDIES: AN INVESTIGATION OF WHY STUDENTS DO NOT RESPOND TO QUESTIONS

Working Paper No. 2003-12 April 2003

Contact: Patricia Dabbs
Field Operations and Publication Management
Patricia.Dabbs@ed.gov

U. S. Department of Education
Institute of Education Sciences

TM035118

BEST COPY AVAILABLE

U.S. Department of Education
Rod Paige
Secretary

Institute of Education Sciences
Grover J. Whitehurst
Director

National Center for Education Statistics
Val Plisko
Associate Commissioner

The National Center for Education Statistics (NCES) is the primary federal entity for collecting, analyzing, and reporting data related to education in the United States and other nations. It fulfills a congressional mandate to collect, collate, analyze, and report full and complete statistics on the condition of education in the United States; conduct and publish reports and specialized analyses of the meaning and significance of such statistics; assist state and local education agencies in improving their statistical systems; and review and report on education activities in foreign countries.

NCES activities are designed to address high priority education data needs; provide consistent, reliable, complete, and accurate indicators of education status and trends; and report timely, useful, and high quality data to the U.S. Department of Education, the Congress, the states, other education policymakers, practitioners, data users, and the general public.

We strive to make our products available in a variety of formats and in language that is appropriate to a variety of audiences. You, as our customer, are the best judge of our success in communicating information effectively. If you have any comments or suggestions about this or any other NCES product or report, we would like to hear from you. Please direct your comments to:

National Center for Education Statistics
Institute of Education Sciences
U.S. Department of Education
1990 K Street NW
Washington, DC 20006–5651

March 2003

The NCES World Wide Web Home Page address is *http://nces.ed.gov*
The NCES World Wide Web Electronic Catalog is: *http://nces.ed.gov/pubsearch*

Suggested Citation

U.S. Department of Education, National Center for Education Statistics. *NAEP Validity Studies: An Investigation of Why Students Do Not Respond to Questions*, NCES 2003–12, by Pamela R. Jakwerth, Frances B. Stancavage, and Ellen D. Reed. Project Officer: Patricia Dabbs. Washington, DC: 2003

For ordering information on this report, write:

U.S. Department of Education
ED Pubs
P.O. Box 1398
Jessup, MD 20794–1398

Or call toll free 1–877–4ED–Pubs

Content Contact:
Patricia Dabbs
(202) 502–7332
Patricia.Dabbs@ed.gov

Foreword

In addition to official NCES publications, NCES staff and individuals commissioned by NCES produce preliminary research reports that include analyses of survey results, and presentations of technical, methodological, and statistical evaluation issues.

The ***Working Paper Series*** was initiated to promote the sharing of the valuable work experience and knowledge reflected in these preliminary reports. These reports are viewed as works in progress, and have not undergone a rigorous review for consistency with NCES Statistical Standards prior to inclusion in the Working Paper Series.

Copies of Working Papers can be downloaded as pdf files from the NCES Electronic Catalog (http://nces.ed.gov/pubsearch/), or contact Sheilah Jupiter at (202) 502–7363, e-mail: sheilah.jupiter@ed.gov, or mail: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, 1990 K Street NW, Room 9048, Washington, DC 20006.

Marilyn M. Seastrom
Chief Mathematical Statistician
Statistical Standards Program

Ralph Lee
Mathematical Statistician
Statistical Standards Program

An Investigation of Why Students Do Not Respond to Questions

Pamela M. Jakwerth
Frances B. Stancavage
Ellen D. Reed
American Institutes for Research

Commissioned by the NAEP Validity Studies (NVS) Panel
March 1999

George W. Bohrnstedt, Panel Chair
Frances B. Stancavage, Project Director

The NAEP Validity Studies Panel was formed by the American Institutes for Research under contract with the National Center for Education Statistics. Points of view or opinions expressed in this paper do not necessarily represent the official positions of the U. S. Department of Education or the American Institutes for Research.

The NAEP Validity Studies (NVS) Panel was formed in 1995 to provide a technical review of NAEP plans and products and to identify technical concerns and promising techniques worthy of further study and research. The members of the panel have been charged with writing focused studies and issue papers on the most salient of the identified issues.

Panel Members:

Albert E. Beaton
Boston College

R. Darrell Bock
University of Chicago

George W. Bohrnstedt, Chair
American Institutes for Research

Audrey Champagne
University at Albany, SUNY

James R. Chromy
Research Triangle Institute

John A. Dossey
Illinois State University

Richard P. Durán
University of California

Larry Hedges
University of Chicago

Gerunda Hughes
Howard University

Richard Jaeger
University of North Carolina

Robert Linn
University of Colorado

Ina V.S. Mullis
Boston College

P. David Pearson
Michigan State University

Lorrie Shepard
University of Colorado

Zollie Stevenson, Jr.
Baltimore City Public Schools

Project Director:

Frances B. Stancavage
American Institutes for Research

Project Officer:

Patricia Dabbs
National Center for Education Statistics

For Information:

NAEP Validity Studies (NVS)
American Institutes for Research
1791 Arastradero Road
PO Box 1113
Palo Alto, CA 94302
Phone: 650/ 493-3550
Fax: 650/ 858-0958

Contents

Introduction 1
Research Questions 3

Methodology 3
Sampling 3
Procedures 4
Interview Protocol and Materials 4
Background Information 5

Results 5
Participation 5
Characteristics of the Student Sample 6
Unanswered Assessment Questions 8
Reasons for Unanswered Questions 9
Other Factors Influencing Nonresponse Rates 12
Implications for Scoring Unanswered Questions 14

Summary 15
Conclusions 15

Recommendations 17
Recommendations for Instrument Development 17
Recommendations forTesting and Test Administration 18
Recommendations for Future Research 19

References 20

Introduction

Over the past decade, developers of the National Assessment of Educational Progress (NAEP) have changed substantially the mix of item types on the NAEP assessments by decreasing the numbers of multiple-choice questions and increasing the numbers of questions requiring short- or extended-constructed responses. These changes have been motivated largely by efforts to encompass the more complex learning outcomes being codified by new curriculum and assessment standards in a number of subject areas. That is, NAEP has attempted to align with widely endorsed recommendations for greater focus on the development and use of higher-order-thinking skills in instruction as well as assessments that better allow students to demonstrate such skills.

With the inclusion of short and extended constructed-response questions on the NAEP assessments, however, researchers have begun to notice unacceptably high student nonresponse rates (Koretz et al. 1993). As a result, NAEP reports, analyses, and subsequent conclusions may be potentially confounded by the fact that large numbers of students are not answering some of the questions. Additionally, nonresponse rates seem to vary with student characteristics like gender and race, which may further impact the validity of NAEP conclusions.

Koretz and his colleagues (1993) conducted an analysis of nonresponse rates on the 1990 NAEP mathematics assessment. They found that, across grade levels, 5 to 10 percent of the items had omit rates of more than 10 percent. The highest omit rates were at grade 12, and almost all of the items with high omit rates were open-ended. Our review of data from recent eighth-grade NAEP assessments in reading and mathematics confirmed the high nonresponse rates associated with some constructed-response items.[1] On average, a given constructed-response item was omitted, or skipped over in the middle of an item block, by about eight percent of these students. In contrast, the average multiple-choice item was omitted by about one percent. The *maximum* omit rates were 18 percent and 4 percent, respectively, for constructed-response and multiple-choice items in the reading assessment and 25 percent and 5 percent, respectively, for constructed-response and multiple-choice items in the mathematics assessment. None of these omit-rate figures include the additional nonresponses of students who stopped short of a given item and failed to answer any further items in that block.

Other researchers (Swinton 1991; Zhu and Thompson 1995) also have found similar overall omit rates across various types of tests. Additionally, they have found that omit rates vary with item and student characteristics and that there are small groups of students for whom omit rates are very high. Of the item characteristics explored in past studies, only format and difficulty seemed to have any significant relationship with the tendency of an item to be skipped. Studies (Koretz et al. 1993; Swinton 1991) have concluded that more open-ended questions tend to be skipped, skipped open-ended questions are often the most difficult, and students seem to stop responding more often at a point where the next question is open-ended rather than multiple-choice.

[1] Specifically, we reviewed grade 8 item responses from the 1992 reading assessment and the 1996 mathematics assessment.

The student characteristics most often examined in relation to high nonresponse rates are gender and ethnicity. Gender most often has been found to be not significant (Koretz et al. 1993; Zhu and Thompson 1995); however, race/ethnicity typically is significant, with Hispanic and African American students omitting answers more often than do Asian and white students (Grandy 1987; Koretz et al. 1993; Swinton 1991; Zhu and Thompson 1995). Reasons for these differences may be related to ability and motivation. Studies have confirmed that some of the differences across ethnic groups can be explained by ability, but

- Grandy (1987) found that 17 percent of high-achieving students still omitted answers, and Koretz et al. (1993) found that, while the percentage of omit rates decreased with higher achievement, it was not by a large factor;
- Differences between ethnic groups were still significant when factoring in achievement (Swinton 1991; Zhu and Thompson 1995);
- Item type still showed a large main effect and had an interaction with ethnicity (Swinton 1991); and
- Some items do have large differences in omit rates across ethnic groups regardless of ability (Swinton 1991; Zhu and Thompson 1995).

In this study, we explored potential reasons behind student omission of responses to assessment questions. Understanding why students fail to answer certain questions may help inform the proper treatment of missing data during the estimation of item parameters and achievement distributions.[2] It may also help test developers identify strategies for increasing response rates for particular types of questions or for particular groups of students.

The study was exploratory, small in scope, and qualitative in nature. The general approach was to visit schools where the 1998 eighth-grade national NAEP assessments in reading and civics were being conducted and interview samples of students about their test taking behaviors and their reasons for not answering particular questions following the assessment sessions. In our interviews we also attempted to determine whether the students could have correctly answered the questions they had left blank. This design was chosen over designs in which the students might take the assessment under more laboratory-like conditions in order to retain the demand characteristics of a typical NAEP assessment. In this way we hoped to obscure as little as possible of the contribution of motivation to NAEP nonresponse.

[2] Current practice is to treat items that are skipped over or omitted in the middle of item blocks as incorrect and to treat unanswered items at the end of blocks as not presented.

Research Questions

Five research questions were investigated. Because of the exploratory nature of the study, we did not expect definitive answers to any of the questions. However, we did hope for insights that could help set directions for future study, including the quantitative analysis of existing NAEP data sets to determine whether observed patterns of association between omissions and student or item characteristics held up over larger numbers of students and items.

Our research questions were as follows:

- What are the reasons students give for not answering questions on the eighth-grade reading and civics NAEP assessments?
- Are students leaving particular types of questions unanswered on these two assessments more often than other types?
- How valid is the assumption that students have attempted, and then passed over, questions left blank in the middle of an item block, but that students have not attempted questions left blank at the end of an item block?
- How valid is the assumption that if students skip items in the middle of an item block, they do not know the answers?
- What modifications can be made to NAEP assessments to decrease the numbers of questions left unanswered?

Methodology

Sampling

Schools were selected from those participating in the eighth-grade 1998 national NAEP assessments. To contain costs, only schools within reasonable driving distance of San Francisco, Los Angeles, or Washington, DC were recruited. In order to maximize the size and diversity of the potential pool of students with unanswered test questions, our initial intent had been to oversample schools that were racially and ethnically diverse and that were anticipated to have significant numbers of lower-performing students. However, because national data collection had already begun by the time the study was launched, the selection of schools was somewhat more limited than originally anticipated. Ultimately, we ended up inviting all public schools within our prescribed geographic areas that tested eighth-grade students during our data collection period. (Three private schools were eliminated from the potential sample because experience with our first pilot site suggested that private school students were less likely than others to leave any questions unanswered.)

One trained AIR staff member visited each participating school on the day of testing and selected students, from the pool of eighth-grade students taking the reading and civics assessments, to participate in debriefing interviews after the testing session. The original intention was to select five students at differing levels of achievement and with differing patterns of nonresponse at each school. However, certain factors complicated the selection process. First, at some schools not all students had parental consent to

participate in the study. This occurred when schools required written permission from parents or, in one case, because the school had only sought permission for a pre-selected subsample of students. Second, at several schools few or no students had omitted any questions. To compensate, more than five students were interviewed at some schools where larger numbers of students met our selection criteria.

Procedures

Prior to the scheduled NAEP administration, schools sent letters to the parents of students participating in the reading or civics assessment informing them of our study. Schools were provided with sample letters and given the choice between letters that asked for active or passive parental consent. We also provided Spanish versions of the letters to one school, and at least one other school sent out its own translation.

In order to make the data collection as nonreactive as possible, schools and students were not told the precise nature of the study, but only that we were interested in how students answer test questions and that we were seeking information that would help improve future test questions. Also, individual students did not know until after the assessment was over whether or not they would be selected to participate in an interview.

On the day of the assessment, the AIR interviewer observed the testing session and then reviewed student test booklets and selected students to participate in the interviews. Selection of students was based on their answer patterns. Ideally, every student interviewed was to have omitted (skipped over in the middle of an item block) at least one question. If fewer than five students at the site had omitted answers to questions, additional students were interviewed, but their data were not used in the final analyses. At sites where more than five students had omitted answers to questions, students were chosen to represent a range in the number and type of questions left unanswered and to represent both the civics and reading assessments.

Test booklets for the selected students were pulled and used in the interviews. Most students were interviewed separately,[3] and interviews were audiotaped. All students participating in the study were given a small gift upon completion of the interview. The majority of the interviews were completed by the project director, and two other people were trained (one in Washington, DC and the other in California) to conduct additional interviews.

Interview Protocol and Materials

Standardized interview protocols were developed. The interview entailed asking students questions about what they thought about particular questions and why they either answered them as they did or did not answer them. The interview protocol was based on protocols developed in AIR's cognitive laboratory for prior studies of student test-taking behaviors. Several staff at AIR reviewed the protocol. A draft version of the protocol was pilot-tested at one of the selected sites, and a second version was pilot-tested at one additional site.

[3] Three students at one school were interviewed as a group to see whether this strategy would elicit new types of information.

Background Information

The background survey that students completed as part of the regular NAEP assessment was used to obtain information about the demographic characteristics of students in the sample. A few of the relevant questions—concerning the highest level of education completed by each parent and the frequency with which either English or a language other than English is spoken at home—were worded slightly different on the reading and civics assessments. However, we recoded the student responses on these questions for comparability.

Data from the interviews were transcribed by each interviewer and entered into a database along with the background variables. The project director coded the interview data for use in subsequent analyses.

Results

Participation

Schools. Sixteen sites participated in the study, including two pilot sites; only one of the pilot sites was included in the analyses.[4] Six of the fifteen valid sites were in northern California, seven were in southern California, and two were in Virginia. Only one other public school in the Washington, DC area (also in Virginia) tested eighth-grade students during the course of this study; however, that site declined to participate. One school in southern California also declined to participate, and one school in northern California had to be excluded because the test date was rescheduled and an interviewer was not available for the new test date.

Ten of the sites used implied consent for the interviews, four sites used signed consent, and at one site the principal pre-identified the students to be interviewed. At one of the implied consent sites, however, only three students took either the civics or reading assessment because the site visit occurred on the day of the NAEP make-up session, which was much smaller than the regular session. Most of the sites were multiracial (predominantly Hispanic and white), and most of them were of low- to mid-socioeconomic standing. Three sites included high proportions of students from extremely impoverished home environments.

Students. Eighty-four students were originally interviewed for this investigation (not including the five students interviewed at the first pilot site). The maximum number of students interviewed at a single site was 10. (In a few cases, we interviewed on two different days, which meant that some students were interviewed one to five days after they had taken the assessment. At one additional site, students were interviewed the day after testing because the interviewer was not available on the day of testing. In all cases, however, interviewers were able to access the students' completed answer booklets and use them during the interviews.)

---

[4] Data from the first pilot site were not included in analyses because the protocol was in an early stage of development, and, furthermore, none of the students available for interviews left any of the questions unanswered. Data from the second pilot site were included because no additional changes were required in the version of the protocol used at that site.

[5] There were some general refusals to take any part of the test.

Characteristics of the Student Sample

Nineteen of the 84 students who were interviewed had no omitted questions. The following analyses, however, focus exclusively on the 65 students who omitted at least one question in the middle of an item block.

The breakdown of this 65-person sample by several demographic variables (i.e., race/ethnicity, gender, mother's and father's level of education, and frequency of speaking another language at home) is presented in table 1. The observed demographics are a function both of the composition of the sampled schools and the characteristics of the students at those schools who omitted questions. In addition, at sites where the potential sample of interviewees (students with omitted questions) was large, the interview sample was chosen to be diverse rather than representative. Consequently, it is not possible to draw statistically meaningful conclusions about the demographic characteristics of students likely to omit questions based on the makeup of our sample. Such questions are better answered by analyzing omit patterns in the full NAEP database.

White and Hispanic students together accounted for over 60 percent of our sample, and males were somewhat more heavily represented than females. With regard to parents' education, the sample was skewed toward those reporting lower or unknown educational attainment. However, one-fifth of the students reported that their mothers had graduated from college and one-quarter reported this level of education for their fathers.

Table 1— Demographic Characteristics of Sampled Students, N=65

	Number and Percent
Race/Ethnicity	
White	16 (25%)
African American	10 (15%)
Hispanic	25 (39%)
Asian	8 (12%)
Other	6 (9%)
Gender	
Male	39 (60%)
Female	26 (40%)
Mother's Education	
Did Not Finish High School	11 (17%)
Graduated From High School	14 (22%)
Some Education After High School	15 (23%)
Graduated From College	13 (20%)
Don't Know	11 (19%)
Father's Education	
Did Not Finish High School	5 (8%)
Graduated From High School	14 (22%)
Some Education After High School	12 (19%)
Graduated From College	17 (26%)
Don't Know	16 (25%)
Other Language	
Never	18 (28%)
Sometimes	25 (39%)
Always	21 (32%)

Over two-thirds of the students reported that they spoke a language other than English at home at least some of the time, and over 25 percent reported always speaking a language other than English at home.

Unanswered Assessment Questions

Numbers of unanswered questions. Reading booklets included either two separately timed reading passages with corresponding questions or one extended reading passage with questions. The shorter passages included 8 to 13 questions and the extended passage included 13 questions. Civics booklets included two separately timed question blocks, with each block containing 18 or 19 questions. Thus, reading booklets contained approximately 20 to 25 test questions (except for booklets containing the extended passage), and civics booklets contained approximately 40 test questions. Reading blocks contained more constructed-response questions than did civics blocks.

As shown in table 2, the number of omitted questions ranged from 1 to 12 for students in this sample who took the reading assessment, and from 1 to 6 for students who took the civics assessment. There was also a wide range in the number of questions not reached (left unanswered at the ends of item blocks). These numbers ranged from 0 to 16 on the reading assessment, and from 0 to 17 on the civics assessment.

Table 2— Average Numbers of Questions Omitted and Not Reached

	Among Sampled Students		
	Mean	***SD***	***Range***
Reading, ~20 test items N=33 students			
Omitted questions	2.8	2.5	1–12
Not-reached questions	3.0	3.8	0–16
Civics, ~40 test items N=32 students			
Omitted questions	2.6	1.2	1–6
Not-reached questions	2.2	3.7	0–17

On average, these students omitted two to three questions and did not reach two to three questions. Numbers of omitted responses and not-reached questions were similar for those taking the reading or civics assessment. However, because the civics assessment booklets contained approximately twice as many questions as the reading assessment booklets, a larger *percentage* of reading than civics questions were left unanswered within each booklet.

Types of specific questions omitted. Almost all omitted questions were short or extended constructed-response questions. Only a few students omitted multiple-choice questions. On the reading assessment, extended constructed-response questions did not seem to be omitted with any more frequency than short constructed-response questions. The civics assessment did not include extended constructed-response questions; however, many of the civics short constructed-response questions, and many of those omitted by higher numbers of students, were questions with two scorable parts (e.g., the student was asked two related questions or was asked to provide two examples of something). For these latter questions, we counted the question as unanswered if either part was left blank.

Students were tested on two blocks of questions each, and unanswered questions came from all assessment blocks. Specific questions were left unanswered by 1 to 8 students for civics questions and 1 to 4 students for reading questions. Nine civics questions and only one reading question were left unanswered by more than three students.

The civics question with the highest omission rate (8 out of the 9 students who left any questions unanswered in that question block) asked a question about the purpose of labor unions. Students who left that question unanswered stated that they did not know what a labor union, was and, when pushed to guess, could not guess correctly. Two other civics questions that were left unanswered by relatively large numbers of students used phrases that students said they did not understand and that probably could have been paraphrased without violating the intent of the question. One of these, left unanswered by 6 of 10 the students who had left any question unanswered within that block, referred to "the democratic process." The other question, left unanswered by 5 of the 6 students leaving questions unanswered within that block, used the phrase "civil disobedience." Many of the students who had skipped over these two questions were able to answer them when the terms were defined. Another civics question, left unanswered by 6 of the 9 students leaving questions unanswered within that block, assessed an understanding of the concept of a constitution for a government.

The reading question that was left unanswered by the most students (4 of 6 who failed to answer at least one of the questions in that block) had to do with the organization of a set of classified ads. One additional reading question that was left unanswered by 3 of 6 students asked students to compare two descriptions of the same character in a story they had read.

Reasons for Unanswered Questions

Lack of knowledge/understanding. Students were questioned about why they left questions unanswered. Of the 65 students in our sample, 30 students (46 percent) indicated omitting at least one of these questions because they understood the question but did not know the answer. Thirty-three (51 percent) indicated omitting at least one of the questions because they either did not understand what the question was asking or they did not understand one or some of the words. One student stated, "These tests confuse me. I don't understand them. I don't understand the sentences and sometimes I don't know what the words mean." Others said,

- "I didn't really get the questions."
- "...well the words, it's kind of like hard to understand."
- "...for most of it, I couldn't figure out what the question was asking."

Students who took the civics assessment were more likely to say they did not know the answer to a question than were students who took the reading assessment. A few students actually had identified correct answers to questions they left unanswered, but they were not sure that they had understood the questions correctly, so they decided not to write the answer.

Missed questions. Six students claimed not to have seen all or part of a question that they had skipped or said that they did not realize they were supposed to answer it. For example, one of the reading questions asked students to fill out a form and two students did not realize it was a question. Three other students said they had their arm over a question and missed it. A few students answered only one part of a two-part constructed-response question (i.e., a question that asked for two reasons, examples, etc.) and stated that they did not realize they had to do both parts. Most students who claimed to have not seen a question were able to answer the question correctly when given the opportunity. Missed questions occurred in both the reading and civics assessments.

Motivation. On the background survey completed as part of the main NAEP assessment, students were asked how important it was for them to do well on the test they took and how hard they had tried on the test compared to other school tests covering similar content. Among the 65 students in our sample, 66 percent of those taking the civics assessment and 73 percent of those taking the reading assessment indicated that doing well on the test was either important or very important. By comparison, across the full national samples for the 1994 and 1992 reading assessments and the 1994 geography and history assessments, approximately 50–55 percent of the eighth-grade students indicated that doing well on the test was important or very important.

Over 80 percent of our sample gave background survey responses indicating that they tried at least as hard as on other tests; this number was similar to those found on other NAEP assessments. During the interview, 63 percent (including 78 percent of the civics and 49 percent of the reading students) said that they would not try harder if the test were graded. Of those students who said that they would try harder, some said that they would take more time reading the questions (and passages), and others said that they would have tried to answer all the questions. A few of the students indicated that they would have studied in advance if the test had been graded.

For most of the students in the sample, therefore, lack of motivation appeared not to be a significant factor. However, eight students did give a specific reason why they did not answer a particular question that indicated a lack of motivation (e.g., they did not have an opinion, they thought it would take too long), and, at two particularly low-income sites, most of the students interviewed indicated being generally unmotivated to answer the test questions. Furthermore, lack of motivation was apparent in the behavior of many other students in these same schools during the testing session (e.g., talking, inattention). Thus, while lack of motivation did not seem to be prevalent in the entire sample of students interviewed, at particular sites—those with the lowest achieving students—it was problematic.

For these students, lack of motivation generally manifested itself in failing to answer constructed-response questions or writing answers such as "I don't care" on the test booklets. Furthermore, while most of these same students would mark answers to multiple-choice questions, they indicated to the interviewers that these were often random marks or guesses.[5]

For example, a student from one of these sites admitted that she did not take the test seriously. She wrote responses like "I don't know and I don't care," and "Hey, whazz-up? I don't know what I am writing because I don't understand nothing." She told the interviewer that she picked her answers to multiple-choice questions by saying "eeny, meany, miny, mo, bubble gum, bubble gum in a dish, how many pieces do you wish, and

you will not be it for the rest of your life, you dirty old dish rag!" The student gave up in the middle of the test, and she did not consider herself a good writer because "I don't know how to express myself." She said that she was talking to her friends during the test. When asked why she did not answer a question that asked her to write a letter, she said that she did not read the article that preceded the question because, "I thought if I read it I wouldn't understand it." She also said that the question itself was not interesting because she did not think she would ever write this type of letter.

The following are examples of the motivation-related reasons for nonresponses given by students at the other sites. One student answered part of a multipart, constructed-response question on the civics assessment but did not answer the other parts because he "got bored." The part of the question that he did answer was correct.

Another student, who took the reading assessment, omitted five questions—all constructed-response. He said that he did not like either of the passages he read and indicated that the first passage, an article that presented a sample of a poet's work, was "odd." For the first question he skipped, the task was to write about images he was left with from the poems. When asked why he did not answer, he said, "No images lingered in my mind." When queried about how he would have answered another (omitted) question if he had been getting a grade on the test, he answered that he would tell them to "ask somebody else." This student was able to answer most of the questions that he omitted when pushed. However, when queried about his preferences among item types, this student indicated that he did not like the extended-response questions because "they put in so much lines, they make you think they expect more...I don't know enough to fill (them up)."

One LEP student who had had particular difficulty with the assessment also admitted to not reading one of the passages because it was too difficult and "there were no pictures and it looked boring." When asked why he did not answer an extended-response question, he replied, "I had to write a really long answer and it seemed like a lot of work." About another question he said that he thought the question was too long and "there are too many words."

Time. Seventy-nine percent of the students in our sample indicated not having enough time to finish the test. Slightly more than half of the students (34) did not reach at least one question. This percentage was around 60 percent for those taking the reading assessment and 40 percent for those taking the civics assessment. Other students indicated skipping over questions and not having enough time to return and finish them. Some of these students read through all of the test questions at least once, but did not have time to return to unanswered questions. Others did not have enough time to reach all of the questions the first time through. This seems to indicate that the tests may be speeded—at least for some groups of students. However, in observations of the testing sessions, interviewers noted that most students finished the test before time was called.

Some students indicated spending a lot of their time on formulating or writing down answers to constructed-response questions, and some indicated spending a lot of time reading either the passage or test questions. Again, there were some responses that were anomalous in this sample. One student ran out of time because he read the passage, then read the questions, and then read the passage again. Another student thought the first constructed-response question was the only question for the section and spent almost all of his time on that question before realizing that there were additional questions. Yet another student indicated being so nervous throughout the testing that he had difficulty concentrating on many of the questions. He also said that he ran out of time because he read all of the questions and was just thinking about them. In fact, when asked what he felt about the test, he replied, "it made me think."

Other Factors Influencing Nonresponse Rates

Test-taking strategy. Test-taking strategies impacted the numbers and locations of questions that students left unanswered. Students who worked through their questions in order without skipping questions tended to have more questions left unanswered at the end of the blocks. Some of these students ran out of time because they spent so much time on an early extended constructed-response question. On the other hand, students who skipped questions did not always return back to them—generally, they ran out of time, but sometimes they decided not to go back.

Most of the students interviewed indicated working through the test questions in order without skipping questions. Students who did skip questions often skipped only those questions to which they did not know the answer.

Students were more apt to skip constructed-response questions. However, only 10 of the students interviewed indicated purposefully skipping constructed-response questions without reading them first. Most students at least read the questions that they skipped, and most said that they read the question at least twice before moving on. Apparently, students were reluctant to attempt an answer to a constructed-response question if they were unsure about the correct answer, and when they perceived that a question would take time to figure out, decided to skip it and go on to other questions, intending to return later.

Over 75 percent of the students indicated guessing on some questions; this percentage was higher among the students who took the civics assessment (88 percent) than reading (67 percent). Close to 40 percent told us that they would always guess on a multiple-choice question, even if they had no idea of the answer and were not able to narrow down the answer choices. Consequently, very few multiple-choice questions were left blank, even on the first pass through the test booklet. Generally, if a student did not know the answer to a multiple-choice question, he/she would guess before going on to another. In sharp contrast, few students presented any evidence of "guessing" (i.e., writing something down when they really did not know the answer) on constructed-response questions.

Testing conditions. The conditions in the testing sessions varied widely. Because civics and writing were tested together, and because the writing assessment sample was the largest of the three, these sessions tended to be larger and to be conducted in larger rooms, such as lunchrooms, multipurpose rooms, or auditoriums, that were less conducive to concentration. Reading sessions, by contrast, tended to be smaller and were conducted in rooms such as libraries that were quieter and had fewer disruptions.

Students were often seated four or five to a table in both the reading and civics sessions. Some schools, however, had the capacity to test smaller groups of students in separate classrooms and to seat students at separate desks. Students in these sessions were quieter than were students in other sessions, and they seemed more likely to answer all their questions.

Conditions at some schools were particularly problematic. At one low-income school, all students were tested together in the same room with separate time being kept for each of the two subjects. At some sites, students were crammed into tables with little room to move their arms. In the sessions where there was a lot of disruption or other factors inhibiting general concentration, it was easier to find students who had left questions unanswered.

One interesting observation was that it was more difficult to find booklets with omitted questions in the schools where the principals had prepared the students ahead of time for the testing session and had stressed the importance of the students doing their best. However, the physical testing conditions also tended to be more favorable in these schools (e.g., smaller rooms, separate desks), and these also were factors found to be associated with higher response rates.

Item formats. Preference for, and past experience with, item format may also impact a student's tendency to leave or not leave a question unanswered. Two-thirds of the students said that they liked multiple-choice questions, and less than one-quarter said that they liked constructed-response questions. Students gave many reasons for their preferences. Reasons for liking multiple-choice questions were:

1) One can guess or get more clues to figure out the answer or understand the question (22 percent);
   - "I like them best because if you don't know you just pick one."
   - "It gives you a hint at the answer and you can check if what you are thinking is wrong."
   - "When I was thinking of the ones I had to write, I really didn't know anything to write so it was kind of easier with the ones that already said something because then you could get a better idea of what it was talking about."
   - "They give you like a choice and you don't have to answer them by yourself."
2) They are easier (15 percent);
   - "They have limited choice."
   - "Easier because you don't have to write so much."
3) One of the answers is always right (14 percent).
   - "Because they give you a couple of choices and one of them has to be right, but on (extended response) we have to guess."
   - "There is always an answer and there is always a chance that you could get it right."
   - "They give you like choices so you don't have to get stressed and find out your own answer."
   - "Because there's a right answer in there and you just have to find it."

Students who said that they liked constructed-response questions usually indicated liking them because they are free to write whatever they want:

- "Because you can write mostly what's in like. . . if the question was like 'how did you feel when so-and-so was um. . .' you could just write anything."
- "You get to explain it in your own words, but if you have only four choices it's pretty hard to pick it out of four."

Seventeen percent of the students said that they did not like answering constructed-response questions because they take too long or made them think too hard:

- "You have to think longer most times."
- "They're like hard to answer because they make you think a lot."
- "Short answers have to come out of your mind."
- "...writing takes too much time."

Other responses included:

- "Has to like have a lot of information."
- "I find them kind of harder because you feel like you have to have the right answer."
- "Worth a lot of points, and if you don't get all the details you lose points."
- "When I was thinking of the ones I had to write, I really didn't know anything to write so it was kind of easier with the ones that already said something because then you could get a better idea of what it was talking about."

Most students felt that they had adequate experience with all question types and said that they get all question types in classes and on tests. Students did not seem to be unfamiliar with the question formats on the test. However, some students said most of the open-ended questions they get are "fill-in-the-blank" or that they usually have only one long essay at the end of a test. A few students also stated that the NAEP tests had more open-ended questions or more writing than they are used to having on one test.

Implications for Scoring Unanswered Questions

In order to have the time to interview five students before the end of the school day, interviews were kept to less than 30 minutes. Therefore, it was not possible to ask students to attempt to answer all unanswered questions. However, approximately two-thirds of the students who had skipped questions could correctly answer (at least for partial credit) at least one of the unanswered questions we queried them about. About one-quarter of the students could do so without any help, seeming only to need time or a little prodding to answer correctly. Others needed questions paraphrased or words defined.

Most students indicated reading the questions they omitted, and many said they read them a couple of times. Most intended to go back to the question(s), but did not have time. By contrast, in most cases, the questions left unanswered at the end of blocks were truly not reached because all but a few students indicated not having the time to even read these questions.

The students who had larger numbers of unanswered questions also seemed to have had difficulty on other questions. However, they tended to guess on the multiple-choice questions and to write something on the open-ended questions that asked for opinions or did not seem to have a right or wrong answer. Many of these students also appeared not to be reading questions carefully.

Summary

Conclusions

What are the reasons students give for not answering questions on the eighth-grade reading and civics NAEP? In general, on these two assessments, students indicated that they left questions unanswered because they either did not have enough time to finish the test or they could not figure out the answer to a particular question. Some students could not figure out the answer to a particular question because they did not understand what the question was asking. However, others understood the question but did not know the answer to the question. The two issues may go hand-in-hand. Students who are less knowledgeable in a specific area may have more difficulty understanding the questions related to that area. Some students were able to answer the question correctly when the question was rephrased. Many students would have been able to get some credit on the question if they had attempted to write something down, but they were reluctant to "guess" on open-ended questions, and it was difficult even to get them to take a guess during the interviews.

The students who had omitted the constructed-response questions most likely would not have done so had they been written as multiple-choice questions. Furthermore, some students would have been able to answer the multiple-choice version of the question correctly because 1) the response options would have clarified the question for some students, and 2) some students knew enough about the question that they probably would have been able to take an educated guess. However, rewording a constructed-response question as a multiple-choice question may change the construct being measured.

Overt lack of motivation to answer the test questions was widespread at only two particularly disadvantaged sites. For many students, however, there seemed to have been an overarching motivation issue—students seemed not to be reading carefully, and many were rushing through the test. The interview responses of some of the students suggested that they might have been more careful if they were being "graded," but self-report on this kind of question may not be a valid indicator of actual behavior.

Are students leaving particular types of questions unanswered on these two assessments more than other types? Just a handful of students left multiple-choice questions unanswered. This is consistent with analysis of recent NAEP mathematics and reading assessments, mentioned in the introduction, which showed that few students were omitting answers to multiple-choice questions. The constructed-response questions that were omitted tended to be longer and more complex questions, and they often were composed of more than one subquestion. Typically, it was a word or phrase in the question that gave a student difficulty.

How valid is the assumption that students have attempted, and then passed over, questions left blank in the middle of an item block, but that students have not attempted questions left blank at the end of an item block? For most students interviewed, questions that were left unanswered in the middle of blocks, technically, should be treated as reached questions (there were very few exceptions of students who did not read these questions at least once). However, it seemed in many cases that if students did not immediately know an answer to a constructed-response question, they skipped it, intending to return to it later if they had time. Thus, they put no real effort into trying to

figure out the answer to the question. Sometimes they did not have time to return back to the question. Students would, on the other hand, put time into trying to figure out the answers to the multiple-choice questions that they could not readily solve. Thus, students seem to be spending less time on constructed-response questions that they cannot immediately answer than they are on multiple-choice questions that they cannot immediately answer. When students feel they are running out of time, they are apt to skip the constructed-response questions.

By contrast, questions left unanswered at the end of blocks of questions were, for the most part, not reached.

How valid is the assumption that if students skip items in the middle of an item block, they do not know the answer? Over half the students could answer at least one of the unanswered questions correctly during the interview. Approximately one-third of the students could do so without any assistance—they just needed time or prodding. Other students needed to have the questions paraphrased or needed to have words defined. Whether this should count against the student is difficult to conclude and must be addressed on a question-by-question basis to determine whether the high reading comprehension level or unfamiliar vocabulary is intrinsic to the construct being measured.

Notably, a large part of the civics test seemed to rely on reading comprehension. For example, students were given a poem and were asked to identify two important ideas the poet was telling others. One student, who did not answer the question said, "I'm not good at poems." Another point to consider vis-à-vis the reading comprehension demands of the civics assessment is that, for good readers, answers to some of the questions could be determined from contextual clues in passages without any prior knowledge. One question, for example, asked what was meant by a term used in a quote.

One could argue that reading ability should not be prerequisite for demonstrating mastery in civics; on the other hand, it may be difficult to address many of the constructs identified in the civics framework without fairly heavy reliance on language. Furthermore, the comprehension problems displayed by many of the students we interviewed went beyond simple decoding; they did not understand the language even when the questions were read to them. Finally, there is a good deal of specialized vocabulary that is arguably intrinsic to the study of civics, and paraphrasing into lay terms may alter the construct being measured.

Language load is a more difficult issue to address in reading. Clearly reading comprehension *is* the construct being measured. However, is it appropriate for part of the reading score to be determined by ability to understand the phrasing or vocabulary used in the questions themselves? For example, some students did not know what the word "constituents" meant in one of the reading questions and skipped the question, but may have been able to answer it correctly, had another word been used. If this happened often and across many students, then we may end up reaching inaccurate conclusions about students' reading abilities.

Recommendations

Recommendations for Instrument Development

- Pay attention to vocabulary. Students had difficulty with several words or phrases that, for the most part, were probably construct irrelevant. Some of these words included: (on reading) constituents, (on civics) democratic process, benefit from, civil disobedience, media. It is important for the test developers to determine if understanding of the particular term or phrase used in the question is essential for demonstrating achievement of the construct. Many students were able to respond to questions correctly once they were paraphrased and words were defined.
- Use simpler phrasing and formatting for the open-ended questions. Many sentences were long and difficult. On the civics test, some students indicated not reading the directions or stimulus text that preceded a question. Instead, they would go straight to the question. Additionally, students seemed to have difficulty with questions that were made up of several subquestions. We suggest that test developers take care to ensure that the reading demands of the questions do not exceed, or are not more difficult than, those of the passage to which they correspond. Twenty-three percent of the students in our sample suggested that the questions be written more clearly so they were easier to understand. Among some of the suggestions were:
  - "I think they could write it a little more clearly and not use such big words."
  - "Some of the words in there are pretty big and you can't understand. . . If they could expand a little bit."
  - "Explain it more like what you mean to give 'em more of questions that you can understand, more smaller words."
  - "Use words that we can actually understand about."
- Add examples. Examples might help with question clarity. Additionally, test developers could consider using some kind of a combination multiple-choice, open-ended item: such as possibly giving students a wrong answer and asking them to supply a correct one or asking students to explain a multiple-choice answer.
- Improve relevance. The suggestions of 11 percent of the students related to improving the reading passages by making them easier or more interesting:
  - "They should put in articles that make kids want to write."

Approximately fifteen percent of the students suggested that the items be more interesting or more related to the lives of the students:

— "Talk more about us. Don't ask questions about things we don't know about."

— "Ask me about what I think we should do about graffiti and I (would) have a lot of ideas."

Recommendations for Testing and Test Administration

- Promote the teaching of test-taking strategies for constructed-response questions. Many students seemed not to know how to attack open-ended questions when they were unsure of the answer or could not understand the question. Students may need training in skills such as breaking questions apart or brainstorming ideas for the answer.
- Consider giving more time or clarify procedures for using time as an accommodation. Ten percent of the students thought they should have more time when asked how the test could be improved. Time seemed to be a factor for most of these students, especially the LEP students. (Schools are required to complete additional background questionnaires for identified LEP students who are eligible for exclusion or for accommodations such as extra time. Several test coordinators indicated that in schools with many LEP students, the schools may choose to have the students tested unaccommodated rather than fill out the supplemental questionnaires. One coordinator stated that he sometimes pulls students himself if he finds they obviously do not understand English. It seems probable that some of the LEP students interviewed—and even some of the non-LEP non-native speakers—could have answered more of the questions, at least with partial credit, if they had had more time.)
- Reconsider the placement of some extended-response questions. Some students who had a long constructed-response question at the beginning of the booklet spent too much time on the question and then ran out of time later. For example, one question asked students to write a letter—of which the format, not the content, would be evaluated for scoring. Some students indicated spending so much time on the content of this letter (which was the third of 12 questions) that they ran out of time before completing the item block.
- Improve and standardize testing conditions. One significant finding in this study was the fact that testing conditions varied widely across sites and that student nonresponse seemed to be related to these conditions. Students were more likely to exhibit behaviors associated with low motivation and were more likely to skip questions under noisier and more crowded conditions. Unfortunately, these conditions occurred more frequently in lower income schools. Thus, it is likely that in the schools that need the most resources, we are obtaining the least valid indications of performance.

Recommendations for Future Research

- Conduct interviews in other subject areas. The percentage of students omitting answers to questions in the testing session observed seemed not to be high. However, in the 1996 NAEP mathematics assessment, omission rates at grade 8 were as high as 25 percent on some questions. Additionally, in mathematics, the extended-response questions were much more complex than were the other two types of questions, and required, to varying degrees, both language and mathematics skills. Thus, it is likely that a replication of this study with the mathematics assessment might yield different results.
- Develop a better picture of the variation in testing conditions and the relationship between performance and testing conditions, when other school and student characteristics are held constant. (Westat test coordinators and quality control monitors currently record some information regarding testing conditions; these data may be sufficient for the proposed analysis.)
- Review student performance in laboratory-like settings. One difficult issue to determine was exactly how much time and effort students devoted to constructed-response questions before leaving them unanswered. Most students said they did read the questions, but it is unclear whether they truly tried to answer them. If there was evidence that students did not try, then treating the questions as incorrect rather than not reached might be questionable. Perhaps, videotaping students in the process of completing the assessment would provide some indication of the effort applied. However, generalizability from the experimental situation to NAEP testing conditions might be low. Another possibility would be having students rate their effort for each question.
- Try out modifications to test forms. Various modifications to test forms could be made to assess the impact on omission rates. One such modification would be to group all extended-response questions together in a separately-timed block, not giving the student the option of skipping them in favor of multiple-choice questions. Another modification would be to embed scaffolding into the questions, which would provide more data on lower achieving students. This last option may be made more feasible with the use of computer adapted testing.

References

Grandy, J. 1987. *Characteristics of examinees who leave questions unanswered on the GRE General Test under rights-only scoring. (GRE Board Professional Report No. 83-16P).* Princeton, NJ: Educational Testing Service.

Koretz, D., Lewis, E., Skewes-Cox, T., and Burstein, L. 1993. *Omitted and not-reached items in mathematics in the 1990 National Assessment of Educational Progress (CRE Technical Report 347).* Los Angeles, CA: Center for Research on Evaluation, Standards, and Student Testing (CRESST).

O'Neil, H. F., Sugrue, B., Abedi, J., Baker, E. L., and Golan, S. 1997. *Final report of experimental studies on motivation and NAEP test performance. (CRE Technical Report 427).* Los Angeles, CA: Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Swinton, S. 1991. *Differential response rates to open-ended and multiple-choice NAEP items by ethnic groups.* Unpublished manuscript, Oct. 23. Princeton, NJ: Educational Testing Service.

Zhu, D., and Thompson, T. D. 1995. *Gender and ethnic differences in tendencies to omit responses on multiple-choice tests using number-right scoring.* (ERIC Document Reproduction Service No. ED 382 689).

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

NOTICE

Reproduction Basis

	This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.
X	This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").